

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) EP 0 899 680 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:  
03.03.1999 Bulletin 1999/09

(51) Int. Cl.<sup>6</sup>: G06K 9/46

(21) Application number: 98202738.5

(22) Date of filing: 15.08.1998

(84) Designated Contracting States:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Designated Extension States:  
AL LT LV MK RO SI

(72) Inventor:  
Luo, Jlebo  
c/o Eastman Kodak Company  
Rochester, New York 14650-2201 (US)

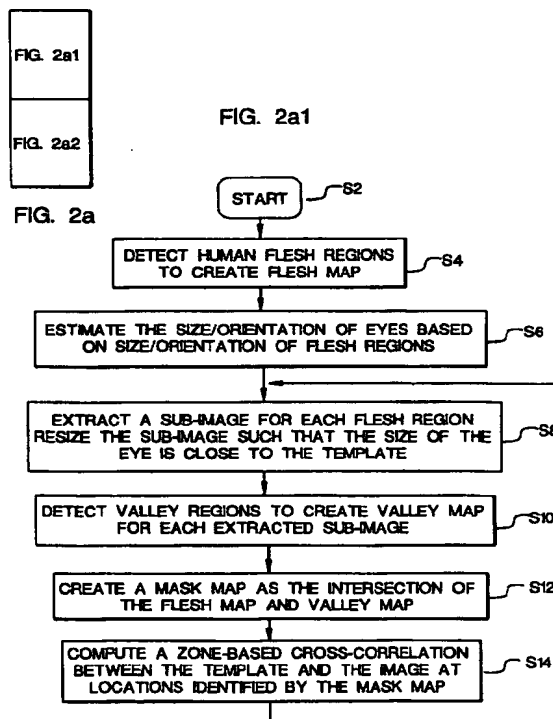
(30) Priority: 29.08.1997 US 920844

(74) Representative:  
Lewandowsky, Klaus, Dipl.-Ing. et al  
Kodak Aktiengesellschaft,  
Patentabteilung  
70323 Stuttgart (DE)

(71) Applicant: EASTMAN KODAK COMPANY  
Rochester, New York 14650 (US)

(54) Method for automatic detection of human eyes in digital images

(57) A computer program product for locating first and second objects, each having substantially the same physical characteristics, and the ratio of the distance between the first and second objects and the size of each object is substantially invariant, the computer program product comprises: a computer readable storage medium having a computer program stored thereon for performing the steps of determining potential flesh regions in an image; determining valley regions in an image; performing template matching for determining a plurality of locations that give a desirable match of the object relative to the template; and performing verification for determining the likelihood of pairs of potential eye candidates at the locations determined in the performing template matching step.



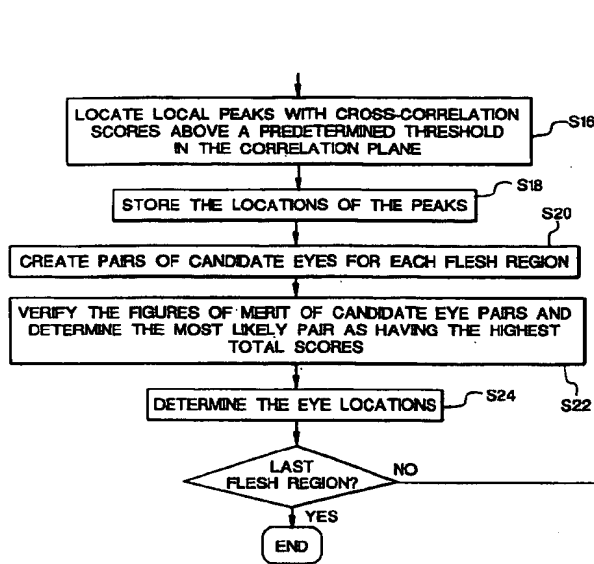


FIG. 2a2

## Description

## APPENDIX

[0001] The disclosure in the appendix of this patent disclosure of this patent document contains material to which a claim of copyright protection is made. The copyright owner has no objection to the facsimile reproduction of any one of the patent documents or the patent disclosure, as it appears in the U.S. Patent and Trademark Office patent file or records, but reserves all other rights whatsoever.

[0002] The invention relates generally to the field of digital image processing and, more particularly, to locating objects in a digital image.

[0003] Identifying objects in an image is performed in a variety of image processing functions. For example, in correcting for red-eye in images, the human eye is located and the undesirable red portion in the eye is replaced with a more aesthetically pleasing color. In the "KODAK" digital print station, the image is displayed on a touch screen and one eye is repeatedly touched for further zooming in on the red portion of the eye upon each touch. The red portion of the eye is then identified by searching for red pixels in the area defined by the zooming process, and the identified red pixels are replaced with a predetermined color for making the image more aesthetically pleasant. The process is then repeated for the other eye.

[0004] A neural networks method of locating human eyes is disclosed in *Learning An Example Selection for Object and Pattern Recognition*, The AI-Lab, MIT by K. K. Sung, November 1995. This method discloses training the a neural net to recognize eyes with acceptable distortion from a pre-selected eye template. The operator repeatedly distorts the original eye template and all variations produced from distorting the eye are labeled as either acceptable or unacceptable. The distorted samples, that is, the training images, and the associated labeling information are fed to the neural net. This training process is repeated until the neural net has achieved satisfactory recognition performance for the training images. The trained neural net effectively has stored possible variations of the eye. Locating an eye is done by feeding a region in the image to the neural net for determining if a desired output, that is, a match, occurs; all matches are identified as an eye.

[0005] Although the presently known and utilized methods of identifying eyes are satisfactory, they are not without drawbacks. The touch screen method requires constant human interaction of repeatedly touching the touch screen for zooming in on the eye and, as a result, is somewhat labor intensive. Still further, the neural net method requires extensive training and is also computationally intensive in the matching process because an exhaustive search has to be performed for all the possible sizes and orientations of the eye.

[0006] Consequently, a need exists for improvements in the method of locating objects in an image so as to overcome the above-described drawbacks.

[0007] The present invention is directed to overcoming one or more of the problems set forth above. Briefly summarized, according to one aspect of the present invention, the invention is directed to a computer program product for locating first and second objects, each having substantially the same physical characteristics, and the ratio of the distance between the first and second objects and the size of each object is substantially invariant, the computer program product comprising: a computer readable storage medium having a computer program stored thereon for performing the steps of: (a) determining potential flesh regions in an image; (b) determining valley regions in an image; (c) performing template matching for determining a plurality of locations that give a desirable match of the object relative to the template; and (d) performing verification for determining the likelihood of pairs of potential eye candidates at the locations determined in step (c).

[0008] It is an object of the present invention to provide a method of finding objects in an image which overcomes the above-described drawbacks.

[0009] It is also an object of the present invention to provide a method of finding objects in an image in an automated manner.

[0010] It is a further object of the present invention to provide a method of estimating the physical size of the objects to be found.

[0011] It is still a further object of the present invention to provide a method of estimating the physical orientation of the objects to be found.

[0012] It is an advantage of the present invention to provide an efficient method of locating objects in an image.

[0013] It is a feature of the present invention to determine an estimated size of each object based on the shape and size of the region where the said objects potential reside.

[0014] It is a feature of the present invention to determine an estimated orientation of each object based on the shape and orientation of the region where the said objects potential reside.

[0015] It is a feature of the present invention to determine a pair (or a group) of objects based on the a plurality of figures of merit determined based on prior knowledge about the relationship between the first and second objects.

[0016] The above and other objects of the present invention will become more apparent when taken in conjunction

with the following description and drawings wherein identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

Fig. 1 is a perspective view of a computer system for implementing the present invention;  
 Fig. 2a is a diagram illustrating the searching procedure used by the present invention;  
 Fig. 2b is a detailed diagram illustrating the zone-based cross-correlation process;  
 Fig. 3 is a detailed diagram illustrating the flesh detection process;  
 Fig. 4 is a detailed diagram illustrating the valley detection process;  
 Fig. 5 is a view of the zone partition of the template by the present invention;  
 Fig. 6 is an illustration of the pairing of eye candidates;  
 Fig. 7 is an illustration of the verification procedure for the distance between and orientations of the two eyes;  
 Fig. 8 is an illustration of matching of the eye-to-eye profile;  
 Fig. 9 is an illustration of the scoring function; and  
 Fig. 10 is an illustration of the face-box and mouth-box.

**[0017]** In the following description, the present invention will be described in the preferred embodiment as a software program. Those skilled in the art will readily recognize that the equivalent of such software may also be constructed in hardware.

**[0018]** Still further, as used herein, computer readable storage medium may comprise, for example; magnetic storage media such as a magnetic disk (such as a floppy disk) or magnetic tape; optical storage media such as an optical disc, optical tape, or machine readable bar code; solid state electronic storage devices such as random access memory (RAM), or read only memory (ROM); or any other physical device or medium employed to store a computer program.

**[0019]** Referring to Fig. 1, there is illustrated a computer system 10 for implementing the present invention. Although the computer system 10 is shown for the purpose of illustrating a preferred embodiment, the present invention is not limited to the computer system 10 shown, but may be used on any electronic processing system. The computer system 10 includes a microprocessor based unit 20 for receiving and processing software programs and for performing other processing functions. A touch screen display 30 is electrically connected to the microprocessor based unit 20 for displaying user related information associated with the software, and for receiving user input via touching the screen. A keyboard 40 is also connected to the microprocessor based unit 20 for permitting a user to input information to the software. As an alternative to using the keyboard 40 for input, a mouse 50 may be used for moving a selector 52 on the display 30 and for selecting an item on which the selector 52 overlays, as is well known in the art.

**[0020]** A compact disk-read only memory (CD-ROM) 55 is connected to the microprocessor based unit 20 for receiving software programs and for providing a means of inputting the software programs and other information to the microprocessor based unit 20 via a compact disk 57, which typically includes a software program. In addition, a floppy disk 61 may also include a software program, and is inserted into the microprocessor based unit 20 for inputting the software program. Still further, the microprocessor based unit 20 may be programmed, as is well known in the art, for storing the software program internally. A printer 56 is connected to the microprocessor based unit 20 for printing a hardcopy of the output of the computer system 10.

**[0021]** Images may also be displayed on the display 30 via a personal computer card (PC card) 62 or, as it was formerly known, a personal computer memory card international association card (PCMCIA card) which contains digitized images electronically embodied the card 62. The PC card 62 is ultimately inserted into the microprocessor based unit 20 for permitting visual display of the image on the display 30.

**[0022]** Referring to Fig. 2a, there is illustrated a flowchart of a software program of the present invention. Before discussing the details of the flowchart, it is instructive to note that, although a portion of the program includes detecting human flesh, any animal flesh may be detected provided the program is modified, as will be apparent to those skilled in the art. The program is initiated S2 and then detection of human flesh is performed to create a flesh map S4.

**[0023]** Referring to Fig. 3, there is illustrated a detailed flowchart of creating the flesh map S4. In this regard, a color image is input S4a into the microprocessor-based unit 20 by any well known means, such as the PC card 62, and is converted S4b into a color space, preferably LST color space. The image code values are then quantized S4c to reduce the total number of histogram bins. A three dimensional (3D) histogram is created S4d for the typically 3-channel color image. This 3D histogram is smoothed S4e to reduce the noise, and the peaks in the 3D histogram are then located S4f. Bin clustering is performed by assigning a peak to each bin of the histogram S4g. For each pixel in the color image, a value is assigned based on the bin that corresponds to the color of the pixel S4h. Connected regions that have at least a minimum number of pixels (*MinNoPixels*), preferably 50 pixels although other values may also be used, in them are labeled S4i. The maximum number of allowed number of regions is *MaxNoRegions* S4i, preferably 20 regions although other values may also be used. Based on the average transformed color component values of human flesh and the average color values of a given region, a flesh probability *Pskin* is calculated for each labeled region S4j. A unique label is assigned to a region with *Pskin* greater than *SkinThreshold* S4k, preferably .7 although other values may also be

used. All the non-flesh pixels are set to zero. If two or more flesh regions touch, they are merged into a single flesh region S4l.

[0024] Referring to Fig. 4, the program receives the flesh map data and performs a well known elliptical fitting technique for each connected flesh region. The aspect ratio and compactness of the fitted ellipse are measured to reject flesh regions of irregular shapes. The aspect ratio is defined as the ratio of the long axis and short axis, and compactness is defined as ratio of the enclosed area by the fitted ellipse and the area of the entire flesh region. If the aspect ratio of the fitted ellipse is above three, it is rejected, or the compactness is below .9, it is rejected. Referring back to Fig. 2a, the program then determines a program estimated size and orientation of the eyes S6 according to the ellipse fitted to each remaining flesh region using the following equation, which is graphically illustrated in Fig. 4:

$$s = b/4$$

where b is the length of the minor axis of the fitted ellipse in pixels and s is the estimate size, or length, of the eye in pixels.

[0025] An estimated angular orientation of the eye is also generated from the orientation of the fitted ellipse S6, as illustrated in Fig. 4. The assumption is that the two eyes are aligned and therefore the orientation of each eye is approximately the same as the orientation of the minor axis of the fitted ellipse. This angle, denoted by  $\theta$ , is between the minor axis and a horizontal line. A sub-image is extracted for each flesh region S8. It is instructive to note that, from this estimated eye size, the resolution of the extracted sub-image is changed so that the eyes in the image have approximately the same size as the eye template S8. As shown in Fig. 6, a particular eye template has a resolution of 19 pixels horizontally and 13 pixels vertically. This resolution change, or resizing, enables the eyes in the images to be matched at the same resolution of a template and against the same amount of structural detail, as will be described in detail herein below. An alternative is to design a set of templates with different amounts of detail and keep the resolution of the image unchanged. Such an alternative design is readily accomplished by those skilled in the art.

[0026] Referring back to Fig. 2a, valley detection is performed to create a valley map for each extracted sub-image S10. The purpose of valley detection is to remove flat flesh regions from further consideration. Referring to Fig. 5, valley detection consists of a plurality of steps. First, a smoothing operation is performed to reduce any noise and small specular highlight in the eye or eyeglasses that may be present, preferably using a morphological opening followed by a morphological closing S10a. The valley regions in the smoothed image are identified, preferably as the output of the difference between a morphologically closed image and itself S10b. The code values of this output image represent the confidence of the presence of a valley. A median filtering operation is applied to the obtained valley image to reduce noise and long thin structures in the image S10c. In general, eyes do not reside in any long thin structure in the valley image. Also, the absolute code value in the valley image does not necessarily correspond to likelihood of an eye socket. However, local maximums in the processed valley image are emphasized, preferably using the difference between the valley image and its morphologically closed version S10d. The result of S10d is combined with the result of texture suppression S10e as well as other inhibitory mechanisms S10f, for example, "redness" of the eye in the case of red-eye detection by a logic AND operation S10g.

[0027] After S10g, pixels with code values greater than a predetermined threshold are set to one, and zero otherwise. A binary valley map is thus created S10h. A mask map for directing the subsequent searching is created as the intersection of the skin map and valley map S12.

[0028] The cross-correlation between the template and the image is computed by sequentially moving the center pixel of the template to each pixel in the searching mask map of each flesh region and performing a specific type of zone-based cross-correlation at each pixel location for determining the center pixel of the eye S14, as will be described in detail below.

[0029] Referring briefly to Fig. 2b, a zone-based cross-correlation S14 is initialized S14a. A template is then retrieved and normalized S14b, if it is not already stored in a normalized state. Referring briefly to Fig. 6, the template is preferably generated from sampling a plurality of eyes and relating their corresponding pixel values, for example by taking the average values at each pixel location. The template is then partitioned into four sub-regions that represent the eyelid, iris, and the two corners of the eye. To normalize the template, the average pixel value for the entire template image is subtracted from each pixel value and the resulting pixel value is divided by the standard deviation of the entire template image for obtaining a normalized pixel value. The resulting template therefore has a mean value of zero and a unit variance.

[0030] More specifically, and referring back to Fig. 2b, with the center of the template at the pixel location of interest, the zone-based cross-correlation includes, first, extracting a block from the image with its center at the current pixel and its size/orientation the same as the template S14c, normalizing the extracted image block S14d, computing the cross-correlation between each sub-region of the extracted block and its counterpart in the template with the pixel of the image at the center of the sub-region S14e, hereinafter referred to as a zone-based correlation. If the cross-correlation for each sub-zone meets or exceeds a predetermined threshold, preferably 0.5, cross-correlation is performed with the

entire template to the same image pixels of interest S14f, hereinafter referred to as a complete correlation. If a threshold, preferably 0.7, is again met, the program temporarily stores the correlation value and the size/orientation of the template in a buffer S14h. If the cross-correlation for one or more sub-zones fails the threshold or the cross-correlation for the entire template fails the threshold, the cross-correlation at the pixel of interest is set to "0" and the associated size/orientation are set to "N/A" S14i. The program then continues to next pixel location in the search mask map S14l for repeating the above-described partitioned and complete correlation operations, if not the last pixel with nonzero mask value in the concerned flesh region.

**[0031]** The above-described zone-based correlation and complete correlation is repeated by varying the template for a plurality of sizes around the estimate size (increasing and decreasing) and a plurality of orientations around the estimate orientation (clockwise and counter-clockwise rotation), in order to refine the size and orientation of the eye S14j. Such increasing and decreasing of the template size/orientation is readily accomplished by those skilled in the art. This refinement involves the same previously described steps, S14c-S14i. If one or more complete correlation scores at a pixel location of interest result in a value above the threshold, the program selects the highest correlation value in the temporary buffer and its corresponding template size/orientation used for obtaining the highest value and places them in memory S14k. It facilitates understanding to note that the above-described varying of the template size is for further refining the estimated size of the eye from Eq. 1, and the size/orientation of the best-matching template variation in turn indicate the exact size/orientation of the actual eye.

**[0032]** For example, the template size is increased by 10% and by 10%. If the highest correlation value is from the 19 x 13 resolution template, the estimated size of the eye is not adjusted. If either of the other resolutions produce the highest correlation value, the estimated size of the eye is adjusted so that it matches the template size producing the highest correlation score. Similarly, the template orientation is increased by 10 degrees and decreased by 10 degrees. If one or more complete correlation scores at the pixel location of interest result in a value above the threshold, the program selects the highest correlation value in the temporary buffer and its corresponding template orientation used for obtaining the highest value and places it in memory. If the highest correlation value is from the template at the original estimated orientation, the estimated orientation of the eye is not adjusted. If either of the other orientations produce the highest correlation value, the estimated orientation of the eye is adjusted so that it matches the template orientation producing the highest correlation value.

**[0033]** As stated previously, the program then continues to next pixel location identified by the search mask for repeating the above-described zone-based and complete correlation S14l after the size and orientation have been refined for the pixel of interest S14k.

**[0034]** The program continues on to verify the most likely candidates from the plurality of peak-correlation points in each window as the center pixel of the eye S16-S24. The peak points are located as the points having a local maximum complete correlation score S16. The locations of these peaks are stored in a buffer S18. Referring to Fig. 6, to verify, a plurality of verification steps are used. The steps involve matching known characteristics about a pair of eyes to all combinations of pixels selected during correlation, and a scoring technique is used (figures-of-merit) to select the most likely pair of locations for the center of the eyes.

**[0035]** Referring both to Figs. 2a and 6, the first step is to form all combinations of pixels selected as likely candidates in the concerned flesh region S20. In other words, each peak pixel is paired with all the other peak pixels in the same flesh region. Referring to Fig. 7, the angular orientation is then determined - the angle between the line formed between the two pixels of interest and a horizontal line through one of the points, preferably the leftwardly pixel. If the angular orientation is not within ten degrees of the estimated angular orientation in S14c, the pair is eliminated as possible candidates for the center of both eyes. If it is within five degrees of the estimated angular orientation, the pair is stored along with its particular score.

**[0036]** Also referring to Fig. 7, the distance between the two candidate eyes is determined. If the distance is not proportional to the size of the eyes according to the knowledge of the human faces, the pair is eliminated as possible candidates for the center of both eyes. If the proportion is within 20% of the normal proportion, the pair is stored along with its particular score.

**[0037]** Referring to Fig. 8, the next step involves taking the pixels along a horizontal line through the two pixels in a possible combination. A graph of code values versus pixel location for each combination will have a shape as illustrated in Fig. 8. If the shape deviates substantially, the pair is eliminated as possible candidates for the center of the eyes; if it does not substantially deviate, the pair is stored along with its particular score. The deviation is preferably determined by the ratio of the middle peak point and the average of the two valley points, although those skilled in the art can determine other suitable measure of the deviation.

**[0038]** Referring to Fig. 10, all combinations are then examined for symmetry. This includes taking the distance between all combinations and, at a distance halfway between them, looking for symmetry on both sides of the image through pixels vertically through this halfway point. The region of interest, which contains the face, preferably has a width of twice the distance between the eyes and a height of three times the distance between the eyes. The face region is divided into two halves - the left side and the right side according to the positions of the eyes. The symmetry is pref-

erably determined by the correlation between the left side and the mirror image of the right side, although those skilled in the art can determine other suitable measure of the symmetry. If symmetry exists for the two sides, the pair and its particular score is again stored; if no symmetry exists, the pair is eliminated as a possible pair of candidates.

[0039] All combinations are also examined for their centrality within the extracted sub-image or the fitted ellipse. A preferred measure of such centrality is defined as the distance between the middle points of the two objects and the major axis of the fitted ellipse, although those skilled in the art can determine other suitable measure of the deviation.

[0040] Also referring to Fig. 10, the image is next examined for the existence of a mouth at an estimated position. The program searches for three or four parallel lines (edges) within a rectangular box that has a width equal to the distance between the eyes and at a predetermined distance from the pair of pixels being analyzed. This distance is 1.2 times the distance between the candidate pairs, although those skilled in the art may determine other distance values or similar criteria. If the lines (edges) exist, the pair and its particular score are stored; if not, the pair is eliminated as possible candidates.

[0041] The combinations are then examined for combined correlation of the two candidates. The combined correlation is the sum of the complete correlation scores at the two candidate locations. If the combined correlation is above a predetermined threshold, the pair and their score are stored; if not, the pair is eliminated as possible candidates.

[0042] The most likely pair is the pair that has the highest cumulative scores S22. The final locations of the eyes are determined by this pair S24. The processes of S8-S24 are repeated for the next identified flesh region until the last flesh region is processed.

[0043] The shape of scoring functions for each above-described figure of merit is illustrated in Fig. 9. With this scoring function, even if a combination fails the threshold of a particular figure of merit, it is assigned a large penalty but can still be retained for further consideration instead of being eliminated as described above. If a figure of merit  $x$  is satisfactory with respect to the threshold  $T_0$ , the output of the scoring function, which is the input to the score accumulator, is close to a normalized maximum value of 1.0. If  $x$  fails the threshold, a increasing amount of penalty is accessed depending on how badly  $x$  fails. The advantage of using such a scoring function is improved robustness if a candidate pair barely fails the threshold but turns out to have the highest cumulative score.

[0044] A computer program written in (whatever language) for performing the steps of the present invention is contained in Appendix A.

[0045] Other features of the invention are included below.

[0046] The computer program product wherein the first and second objects are first and second human eyes and wherein computing the proportion includes computing a distance between the first and second eyes for satisfying an anthropological prior model.

[0047] The computer program product wherein the first and second objects are first and second human eyes and wherein computing the profile includes predetermining a model of an eye-to-eye profile and determining an actual profile from the image and computing a goodness-of-fit between the actual profile and the model profile.

[0048] The computer program product wherein the first and second objects are first and second human eyes and wherein comprising computing the symmetry includes computing the symmetry between first and second halves of a face window determined by the eye locations.

[0049] The computer program product wherein the first and second objects are first and second human eyes and wherein computing the evidence of mouth includes computing strength and orientation of edges within a mouth window determined by the eye locations.

[0050] The computer program product wherein the first and second objects are first and second human eyes and wherein computing the centrality includes computing a distance from a middle point between the first and second eyes to a major axis of the determined flesh region.

[0051] The computer program product wherein the first and second objects are first and second eyes and wherein computing a combined correlation score includes summing up the individual correlation scores obtained from the matching obtained in step (g).

[0052] The computer program product further comprising the step of normalizing the template for maximizing robustness of matching.

[0053] The computer program product further comprising the step of creating an image block of a same size and orientation as the template.

[0054] The computer program product further comprising the step of normalizing the image block for maximizing robustness of matching.

[0055] The computer program product further comprising the step of extracting individual sub-zones in both the template and the image block, and performing cross-correlation for each corresponding sub-zone.

[0056] The computer program product further comprising the step of computing an overall cross-correlation score for the image block if the cross-correlation score for each sub-zone exceeds a threshold.

**PARTS LIST**

**[0057]**

5	10	computer system
	20	microprocessor-based unit
	30	display
	40	keyboard
	50	mouse
	52	selector
10	55	CD-ROM
	56	printer
	57	compact disk
	61	floppy disk
	62	PC card
15		

20

25

30

35

40

45

50

55



## APPENDIX

Page

1

```

5
10
15
20
25
30
35
40
45
50
55

/******
/* Name: eyeivs.cc
/* Function: semi-automatic eye location based on template matching
/* Feature: . using two eye search windows determined by initial inputs to limit the scope of searching
/*          . or using default eye search boxes if images are in standard format
/*          . using input of the approximate two eye locations (user input initial positions)
/*          . approximate scale is estimated
/*          . approximate tilt is estimated
/*          . no other prescreening (morphological operations)
/*          . one winner
/* Author: Jiebo Luo
/* Copyright: (c) Eastman Kodak Company, 1997
/******/

/* standard header files */
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>

/* customer header files */
#include "image.h" // header file of C++ image processing library

/* customer default right-eye box for passport style pictures */
#define BOXHEIGHT1 0.3
#define BOXWIDTH1 0.3
#define BOXLEFTOFFSET1 0.1
#define BOXABOVEOFFSET1 0.1

/* customer default two-eye boxes for passport style pictures */
#define BOXHEIGHT2 0.1
#define BOXWIDTH2 0.1
#define ABSBOXHEIGHT2 21 /* was 11 */
#define ABSBOXWIDTH2 21 /* was 11 */
#define DRIFTAWAY 10 /* was 5 */

/* customer function definition */
#define nint(x) (x)>0 ? ((int)((x)+0.5)) : ((int)((x)-0.5)) /* always wrapped */
#define lmax(x, y) (x)>(y) ? (x) : (y)
#define lmin(x, y) (x)<(y) ? (x) : (y)
#define lsqr(x) ((x)*(x))
#define pvalid(x) ((x) >= 0 && x <= 255) ? 1 : 0
#define TRUE 1
#define FALSE 0
#define SETPEAK 1

/* customer constant definition for correlation methods */
#define INTENSITY 1
#define NORMINTENSITY 2
#define GRADIENT_ABS 3
#define GRADIENT_POL 4
#define LAPLACIAN 5
#define PHASE_ONLY 6 /* Phase-Only Filter */
#define BIN_PHASE_ONLY 7 /* Binary Phase-Only Filter */

/* customer constant definition for scale/tilt variation */
#define MINSCALE 0.55

```

```

5
10
15
20
25
30
35
40
45
50
55

#define MAXSCALE 1.45
#define STEPSCALE 0.15
#define MINTILT -10.0
#define MAXTILT +10.0
#define STEPTILT 5.0

/* customer definition of goodness for eye subregions */
#define FUZZYNEIGHBORHOOD 1
#define PEAKNEIGHBORHOOD 2

/* customer definition of eye-searching window */
#define SCOPECENTER1 /* default search window geometrically centered */
#define SCOPEWIDTH1 0.4 /* 40% in total horizontally */
#define SCOPEHEIGHT1 0.2 /* 20% in total vertically */
#define SCOPECENTER2 /* manual search window (smaller) */
#define SCOPEWIDTH2 0.2 /* 20% in total horizontally */
#define SCOPEHEIGHT2 0.1 /* 10% in total vertically */

/* external functions */
image morpho(image &a, int OPERATION, int ELEMENT_TYPE, int ELEMENT_SIZE);
/* perform a morphological operation specified by OPERATION using an element specified by ELEMENT_TYPE and ELEMENT_SIZE
image median(image &a, int KERNEL_TYPE, int KERNEL_SIZE);
/* perform median filtering using the specified kernel
image texturemask(image &a);
/* generate a map of texture activities
int skindetect(image &a, image &sa);
/* skin detection
image regioncrop(image &a, image &sa, int REGION);
/* crop a rectangular image block that includes the skin region
float ellipsefit(image &ak, int &minor, int &major, float &angle, float &aspect);
/* perform elliptic fitting and returns ellipticity measure
image masking(image &ak, image &sa);
/* create searching mask map using a combination of morphological operations, texture masking for a skin region
int range(int idim, int jdim, int i, int j);
/* check if a pixel is within the range of the image/searching-box
image mcorrelator(image &a, image &templ, image &mask, int maskthresh, int mflag);
/* compute the zone-based cross-correlation with the template
image scale(image &a, float scalex, float scaley);
/* perform scaling of an image using bilinear interpolation
void update(image &aaa, image &aa, float s);
/* update the standing correlation map with the current correlation map
image eyepartition(int nrows, int ncols);
/* partition the eye into sub-zones
int peak(image &a, float threshold, int wsize, int setpeakflag);
/* locate the local peaks in the correlation map
float maskvariance(image &a, image &mask, float mean);
/* calculate the variance of a masked region
farray fmedian(farray &x, int width);
/* perform median filtering of an farray
farray faverage(farray &x, int width, int norm);
/* perform running-average filtering of an farray
void boxmask(image &mask, int x1, int xj1, int xj2, int xj2, int boxwidth, int boxheight, int leftflag, int rightflag);
/* determine the eye boxes

```

Using for J1200 Line

Date: Jan 24 09:35:00 1997

Page

3

```

5
    image lum(image &a);
    // convert the image into luminance signals
    void pairing(image &im, PEAK peaks[], int pnun, float templwidth, int halfheight,
10    int halfwidth, float stepscale, float steptilt, int &kleft, int &kright, int
    x1l, int xj1, int xi2, int xj2, int boxwidth, int boxheight);
    // form and evaluate pairing of two eye candidates
    void marking(image &a, int il, int jl, int ir, int jr);
    // mark the final eye locations
    int get_ee_line(image &a, int x1, int y1, int x2, int y2, farray &profile, int
15    DO_DRAW);
    // get an intensity profile along the eye-to-eye line
    float wprofile(farray &x, int &lefteye, int &noseridge, int &righteye);
    // measure the goodness of the eye-to-eye profile expected to be W-shaped
    float mouth(image &a, int x1, int y1, int x2, int y2, int &xm, int &ym, int GFLA
    G);
    // find evidences of mouth

    /* Default Threshold Values for Zone-based Matching */
    float GOODEYEBROW      = 0.25;
    float GOODEYECORNER    = 0.25;
    float GOODEYEIRIS      = 0.25;
    float GOODPEAKCOR      = 0.5;
    float GOODPEAKCOR2     = 0.70;

    class PEAK {
    public:
25        int peaknum;
        int i, j; /* position */
        float s, t; /* scale and tilt */
        float score; /* correlation score */
    }; /* define an object for candidate peaks */

    image epart;

    float t_global, s_global;

30    /* begin main() */
    int main(int argc, char **argv)
    {

        // parse a command line

35        // declaration of variables and objects
        image a, sa, ak;

        x1l0 = x1l; xj10 = xj1; xi20 = xi2; xj20 = xj2;
        // back-up initial positions

        boxaccept = 1;
        a.read(imginname, short_);
        // load the original image

40        k = skindetect(a, sa);
        // skin detection routine returns the total number of skin regions

        for(region = 1; region <= k; k++)
        {
45            ak = regioncrop(a, sa, region);
            // crop the k-th flesh reigon

```

Listing for J1560 L10

The July 24, 1985 Edition

Page

4

```

5      ellipticity = ellipsefit(ak, minor, major, angle, aspect);
        // elliptical fitting

10     if ( ellipticity < MIN_ELLIPTICITY) continue;

        // resize the cropped image according to the estimated eye size */
        if(ak.chan() == 3) ak = lum(ak);
        // convert the image to luminance signals

15     mask = masking(ak, sa);
        // create searching mask map using a combination of morphological operations, texture masking

        epart = eyepartition((templ.rows()),(templ.cols()));
        // partition the eye template

        // determine a range of sizes and orientations for fine matching
20     // begin correlation for matching

    for(float s=maxscale; s>=minscale; s-=stepscale)
    {
        for(float t=mintilt; t<=maxtilt; t+=steptilt)
        {
25             aa = mcorrelator(aa,templ,mask,maskthreshold,mflag);

            printf("/* peak detection in the transformed correlation plane*\n");
            pnum = peak(aa, GOODPEAKCOR*127, PEAKNEIGHBORHOOD, 0); /*half window*/
            printf("Totally %d peaks found at s=%f, t=%f.\n", pnum, s, t);

            printf("/* peak detection in the inversely transformed correlation plane */\n");
            pnum = peak(aa, GOODPEAKCOR*127, PEAKNEIGHBORHOOD, 0); /*half window*/
30             printf("Totally %d peaks found at s=%f, t=%f.\n", pnum, s, t);

            if(s == maxscale && t == mintilt) aaa = aa.copy(float_);
            else update(aaa, aa, s); /* copy/update the final correlation map */
        }
    }

35     printf("\n* Checking Local Peaks in the S-T Composite Correlation Map *\n");
    pnum = peak(aaa, GOODPEAKCOR*127, PEAKNEIGHBORHOOD, SETPEAK); /*half window*/
    printf("Totally %d peaks (255) found at all scales and tilts.\n", pnum);

    int kleft=1000, kright=1000;
    int plen;
    farray profile(256), profile0, profile1, profile2;

40     a0 = ak.copy(byte_);
    pairing2(a0, peaks, pnum, (float)templwidth, halfheight, halfwidth, stepscale,
    steptilt, kleft, kright, x11, xj1, xi2, xj2, boxwidth, boxheight);

    printf("Marking eye locations: left @(%d,%d) right @(%d,%d) ...\n", x11, xj1, xi2, xj2);

45     plen = get_ee_line(a0, x11, xj1, xi2, xj2, profile, 1);
    marking(a0, x11, xj1, xi2, xj2);

```

Listing for Jiebo Luo

Thu Jul 24 09:56:38 1997

Page

5

```

5
    profile0.alloc(plen);
    for(i=0;i<plen;i++)
        profile0.ptr[i] = profile.ptr[i]; //printf("%d\t", (int)profile[i]);)
10
    printf("Eye-to-eye profile length = %d\n", plen);
    profile.write("profile.dat");
    printf("Profile Filtering: window width = %d\n", plen/16*2+1);

    profile1 = profile0.median( plen/16*2+1 );
    profile2 = faverage(profile1,plen/16*2+1, 1);
15
    int lefteye, noseridge, righteye;
    goodwp = wprofile(profile1, &lefteye, &noseridge, &righteye);
    printf("goodness of the E2E profile = %f\n", wprofile(profile2, lefteye, no
seridge, righteye));

    mouth(a0, x11, xj1, x12, xj2, im, jm, gflag); /* detect and mark the mouth
*/
20
    )

    printf("compensate for subsampling factor = %d\n", fac);
    x11 = x11*fac;
    xj1 = xj1*fac;
    x12 = x12*fac;
    xj2 = xj2*fac;
25
    a0.write(imgoutname, byte_);
    print("Processed one skin region.\n");
    )

    printf("Process done. \n");

    mask.dealloc(); mask0.dealloc(); mask1.dealloc();
    profile.dealloc();profile0.dealloc();profile1.dealloc();profile2.dealloc();
30
    // free memory space
} // end main()

```

# Claims

1. A computer program product for locating first and second objects each having substantially the same physical characteristics, and the ratio of the distance between the first and second objects and the size of each object is substantially invariant, the computer program product comprising: a computer readable storage medium having a computer program stored thereon for performing the steps of:
  - (a) determining a potential flesh region in an image
  - (b) determining valley regions in an image;
  - (c) performing template matching for determining a plurality of locations that give a desirable match of the object relative to the template; and
  - (d) performing verification for determining the likelihood of a pair of potential object candidates at the locations determined in step (c).
2. The computer program product as in claim 1 further comprising either or both (e) determining an estimate size of both the first and second objects based on a shape and size of a determined flesh region or (f) determining an estimate orientation of both the first and second objects based on the shape and orientation of the determined flesh region.
3. The computer program product as in claim 1 further comprising forming a mask image for searching for the first and second objects; the locations where a search is to be performed are determined by the valley regions within the determined flesh region.
4. The computer program product as in claim 1 further comprising (g) reiteratively positioning a template on the locations determined by the mask image for determining a location that gives a desirable match of the object relative to the template.
5. The computer program product as in claim 4 further comprising the step of identifying desirable locations that give locally maximum cross-correlation score.
6. The computer program product as in claim 5 further comprising forming a pair of objects from locations with desirable matching response to the template and within the determined flesh region; and determining desirability of each hypothesized pair.
7. The computer program product as in claim 6 further comprising finding a best pair of locations from all possible combinations of identified locations of both the first and second objects.
8. The computer program product as in claim 7, wherein the step of finding best pair of locations includes computing a plurality of figures of merit and determining a pair of locations that give a best combined figure of merit.
9. The computer program product as in claim 8, wherein computing a plurality of figures of merit includes computing individually or in combination an orientation, proportion, profile, symmetry, evidence of mouth, centrality, or combined correlation score.
10. The computer program product as in claim 9, wherein the first and second objects are first and second human eyes and wherein computing the orientation includes measuring the difference between an orientation of a line connecting the first and second eyes, and an average orientation of the first and second eyes.

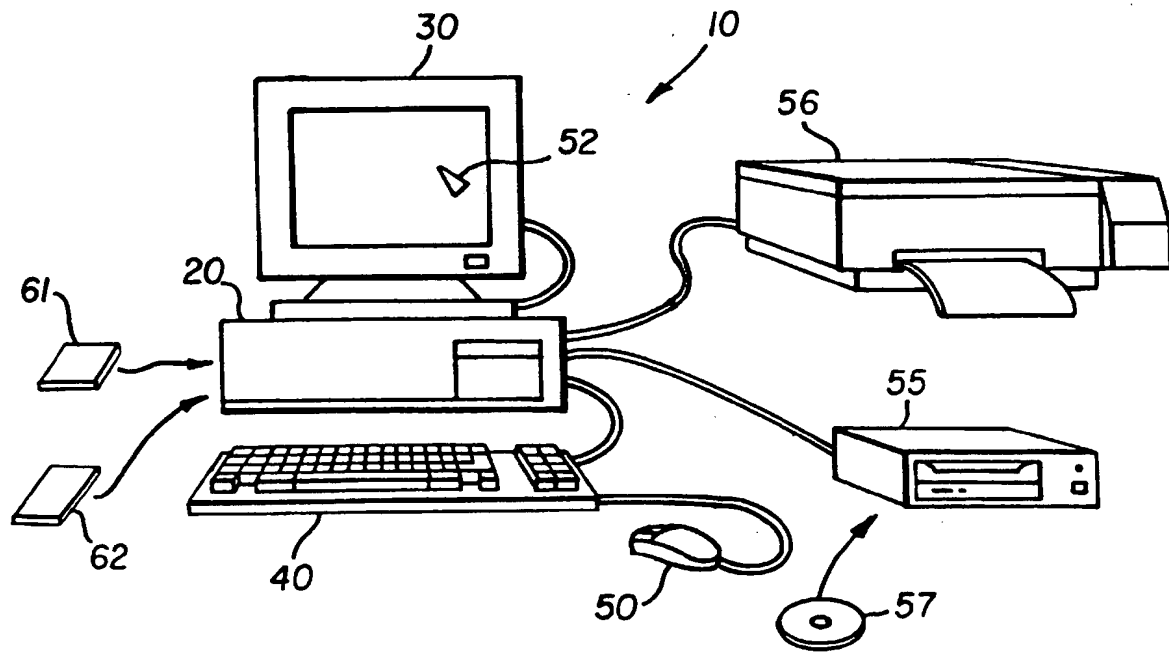


FIG. 1

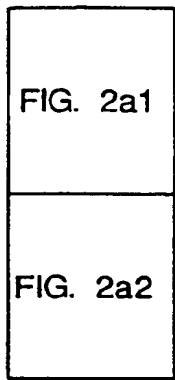
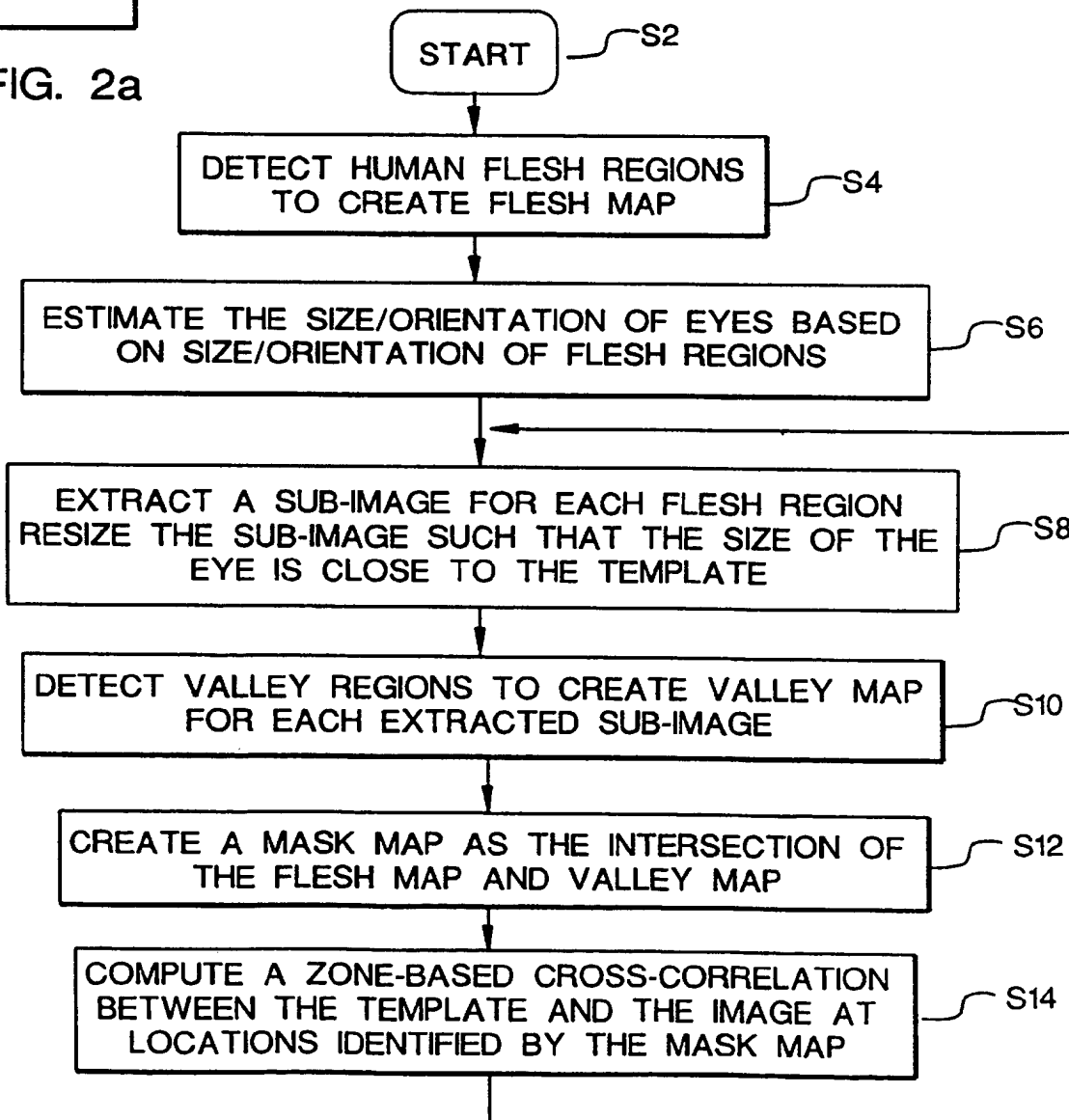


FIG. 2a

FIG. 2a1





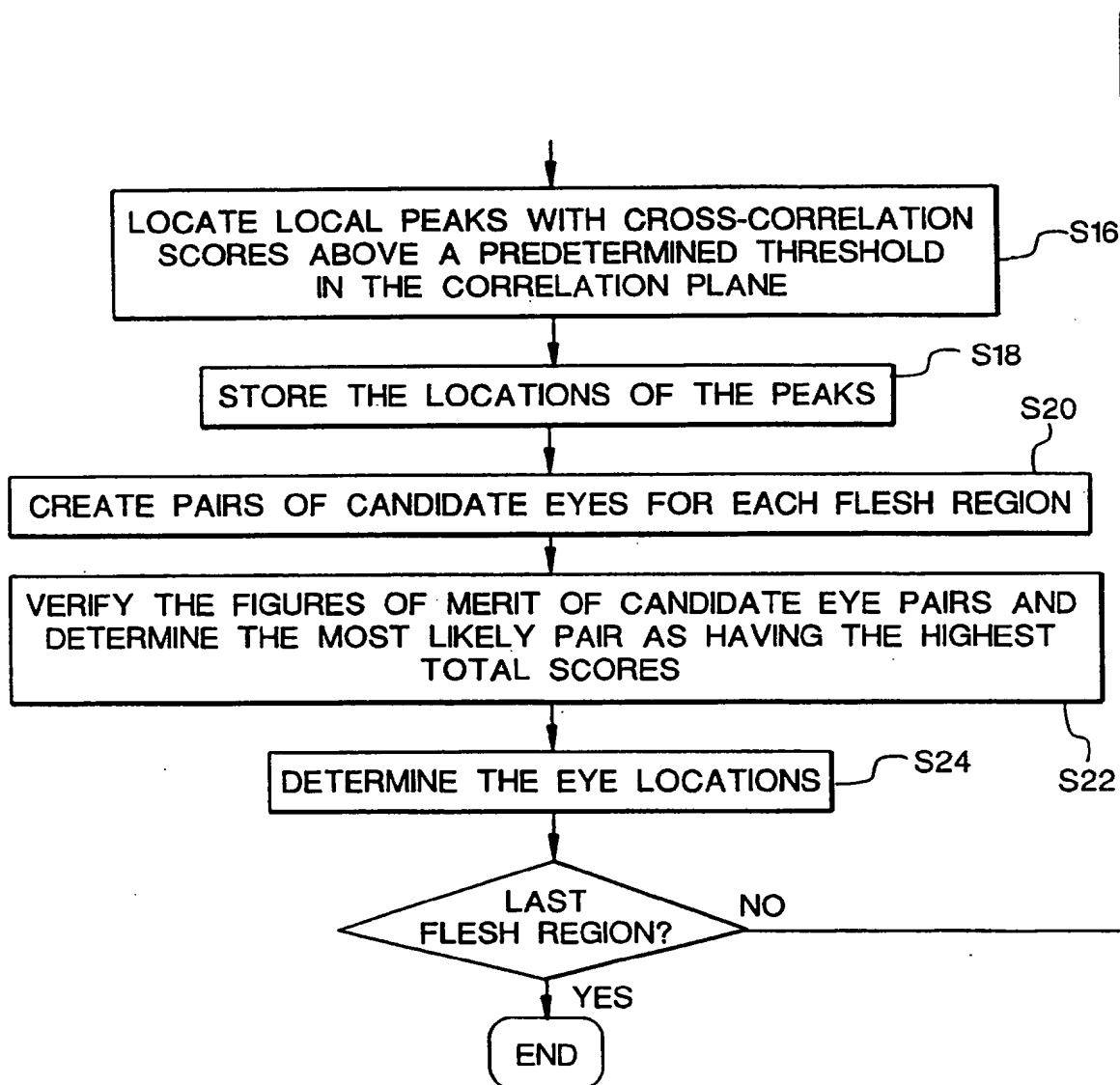


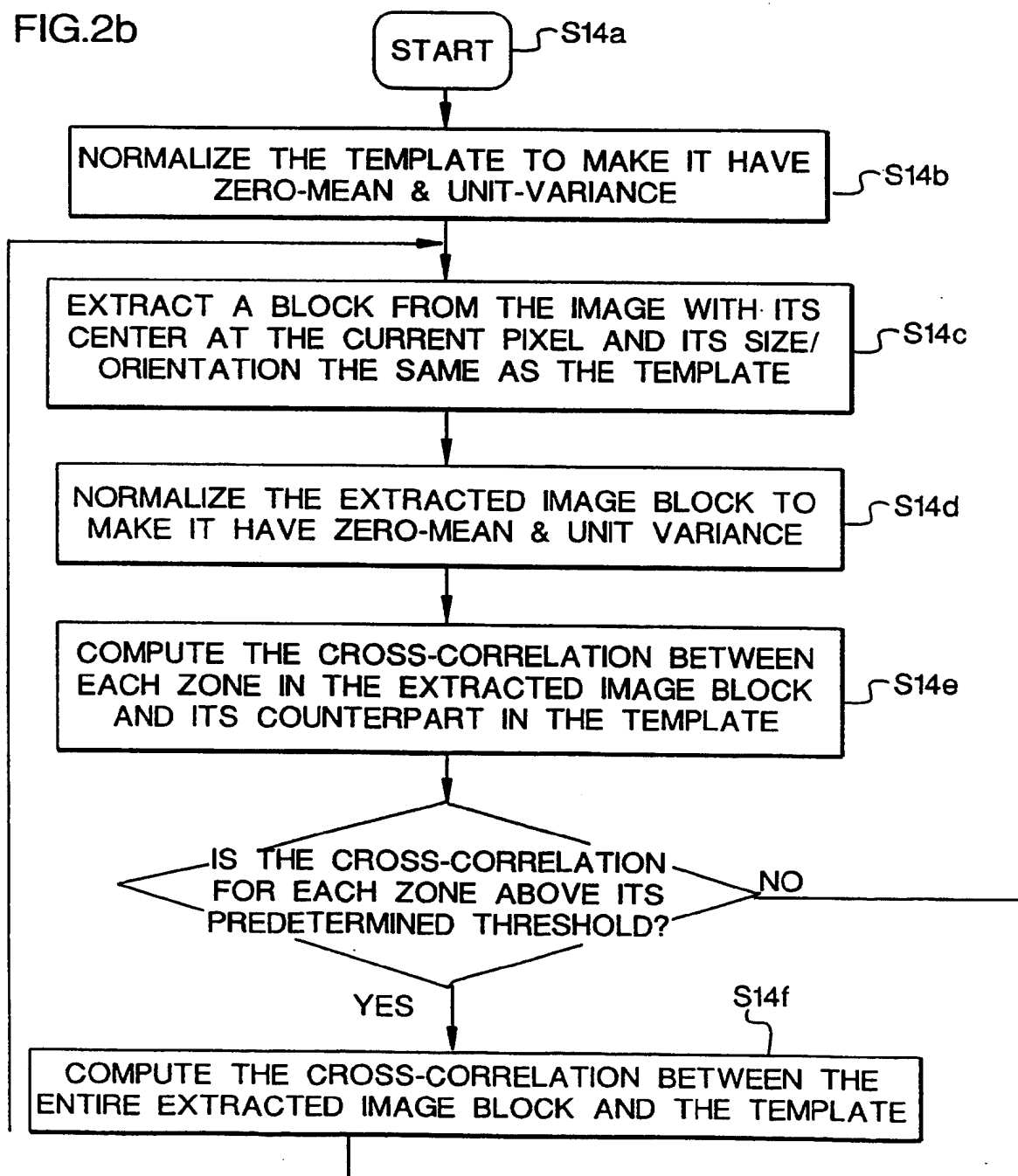
FIG. 2a2

FIG.2b1

FIG.2b2

FIG. 2b1

FIG.2b



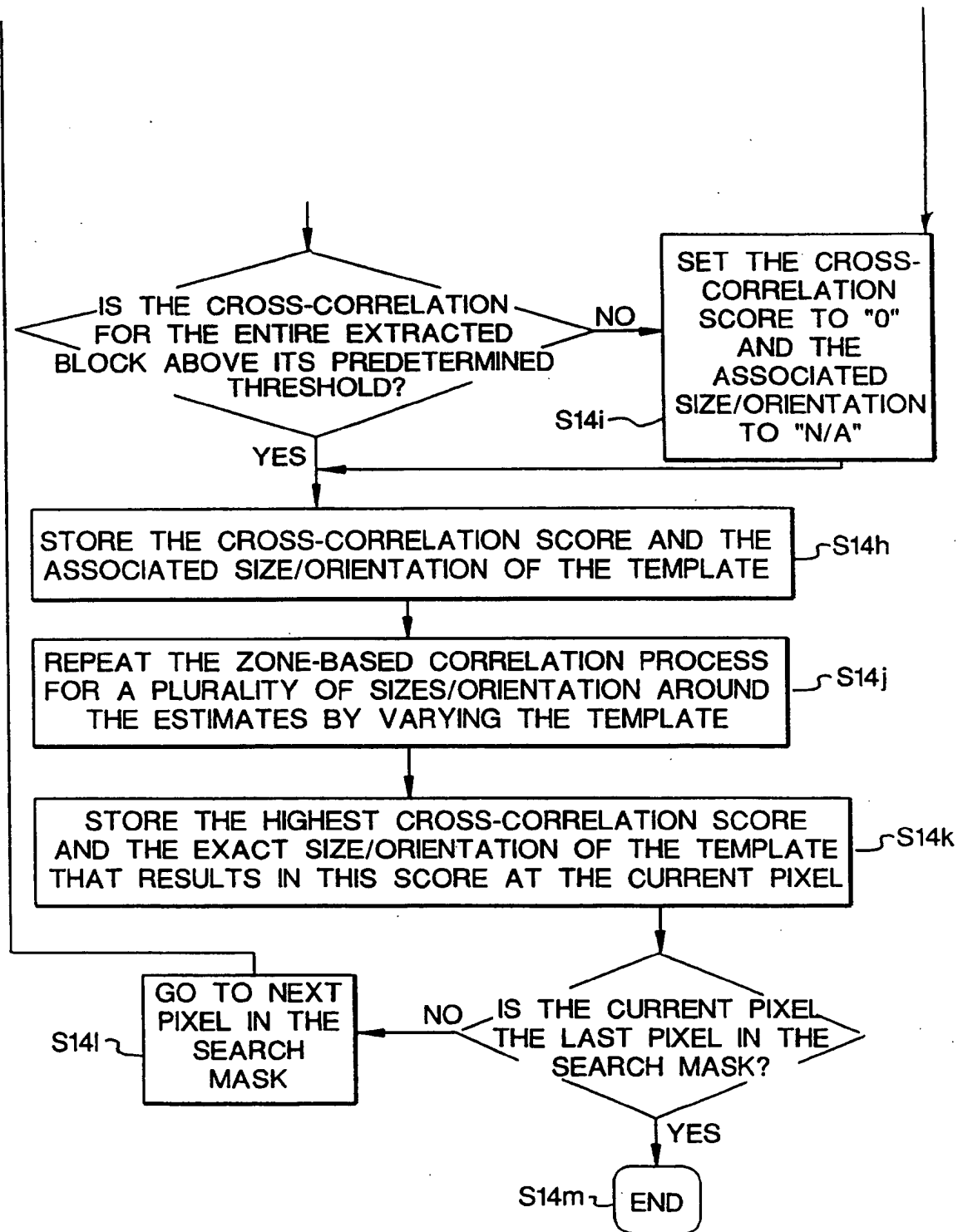


FIG. 2B2

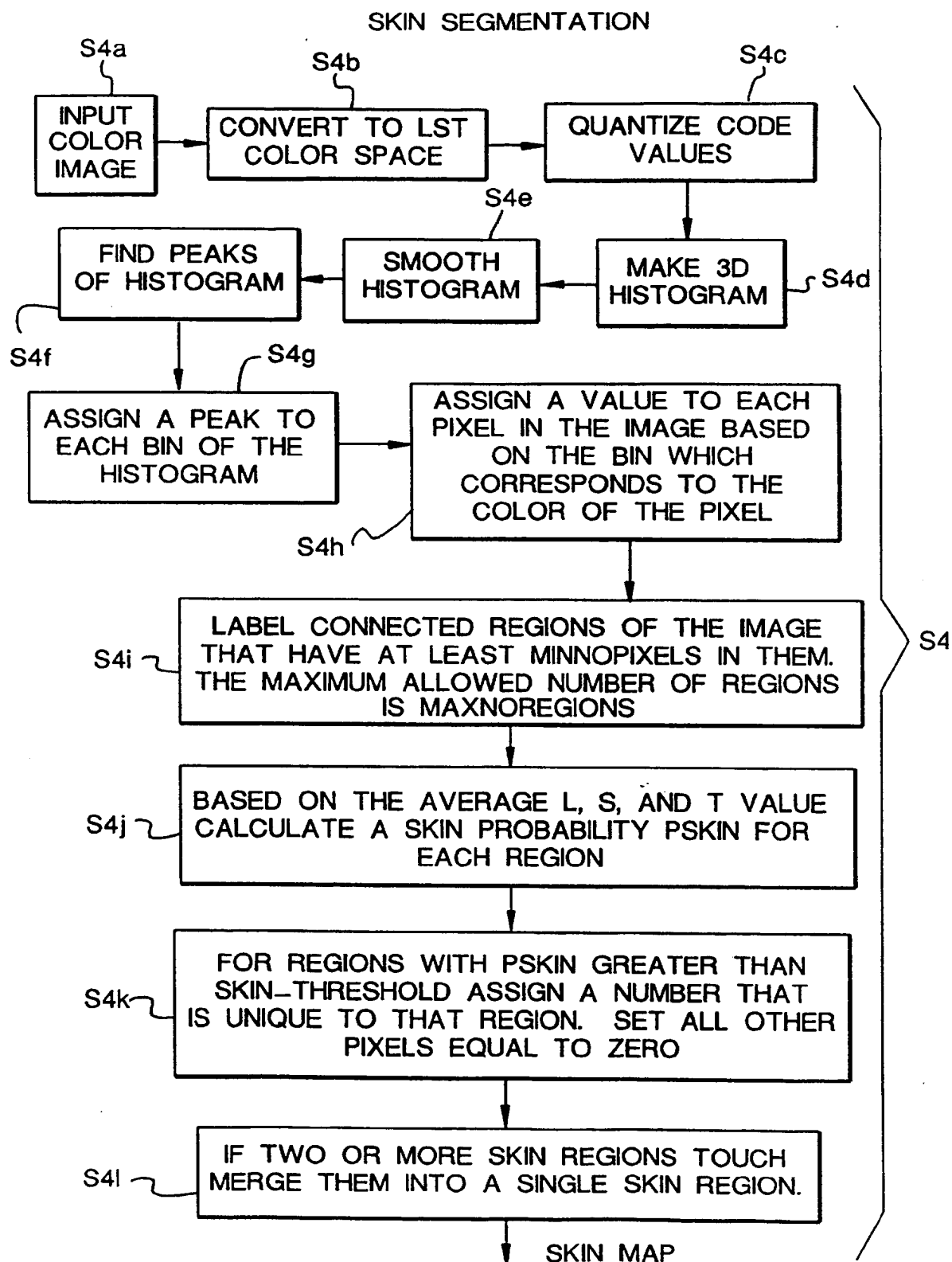


FIG. 3

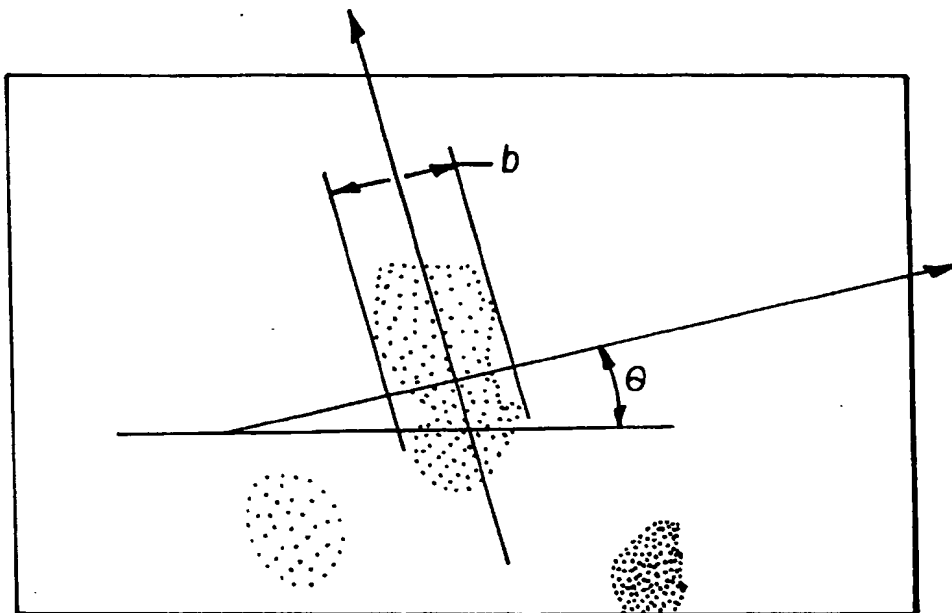
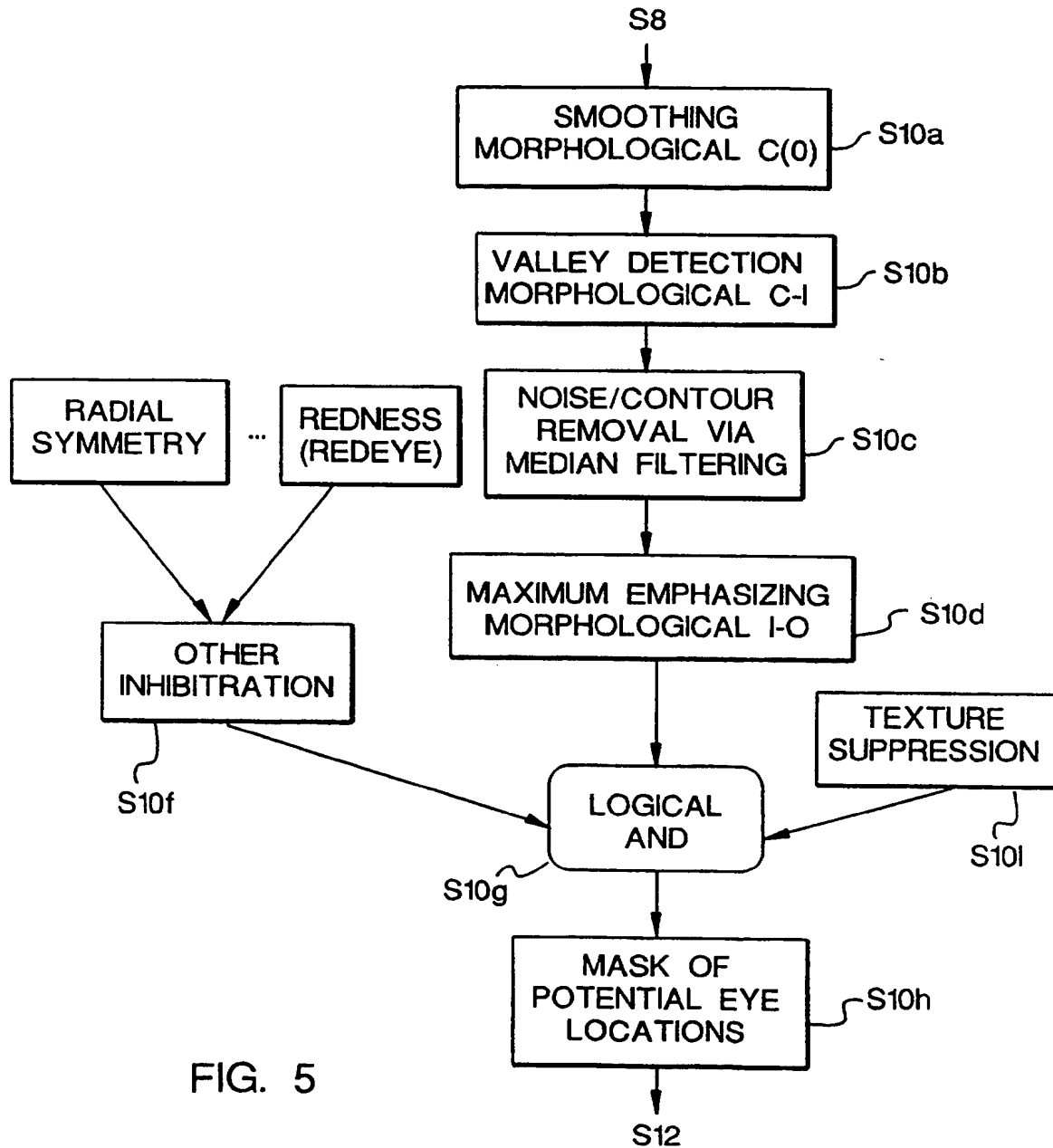
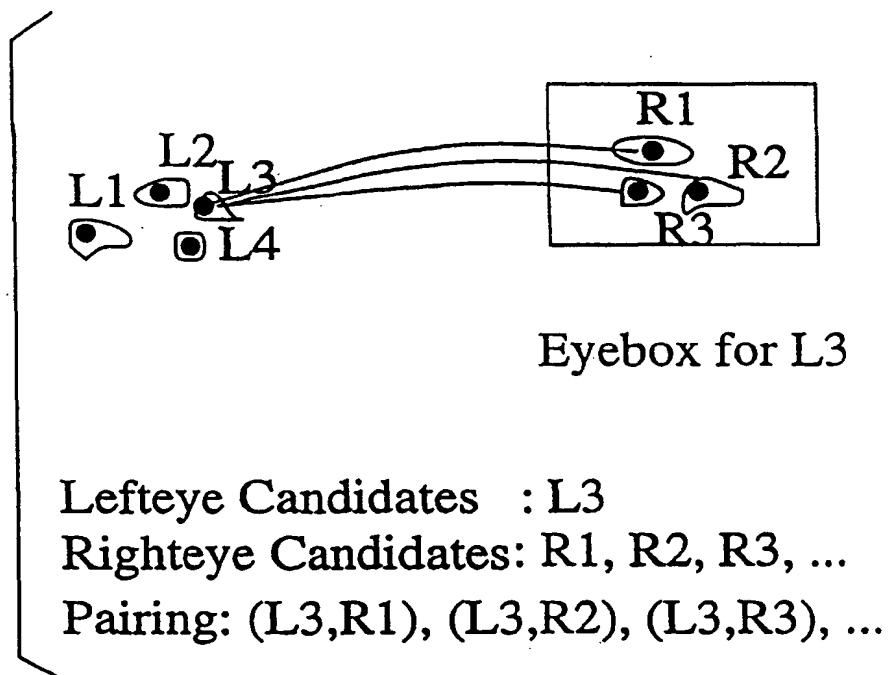
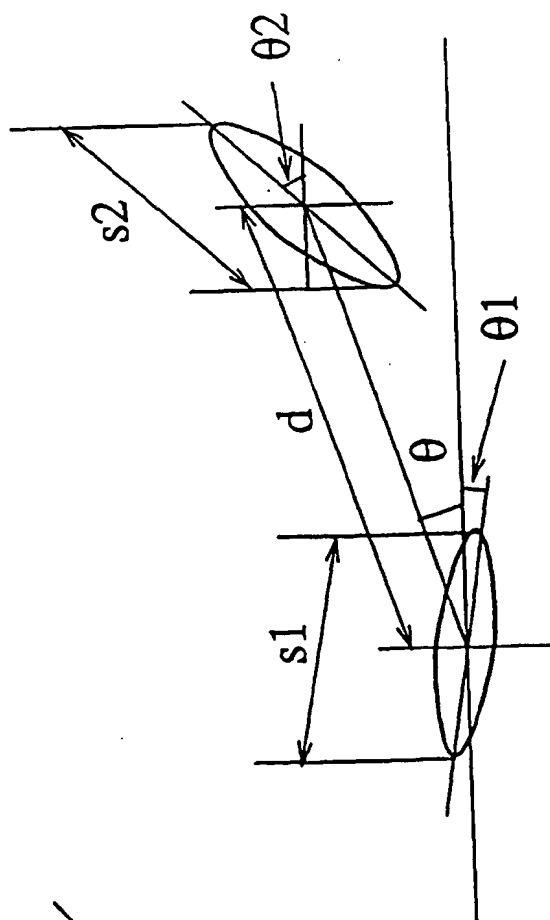


FIG. 4





**FIG. 6**



- $\theta_1$ : orientation of the first object (in this case negative)
- $\theta_2$ : orientation of the second object (in this case positive)
- $\theta$ : orientation of the line connecting the centers of the first and second objects
- $d$ : distance between the centers of the first and second objects
- $s_1$ : size of the first object
- $s_2$ : size of the second object

FIG. 7



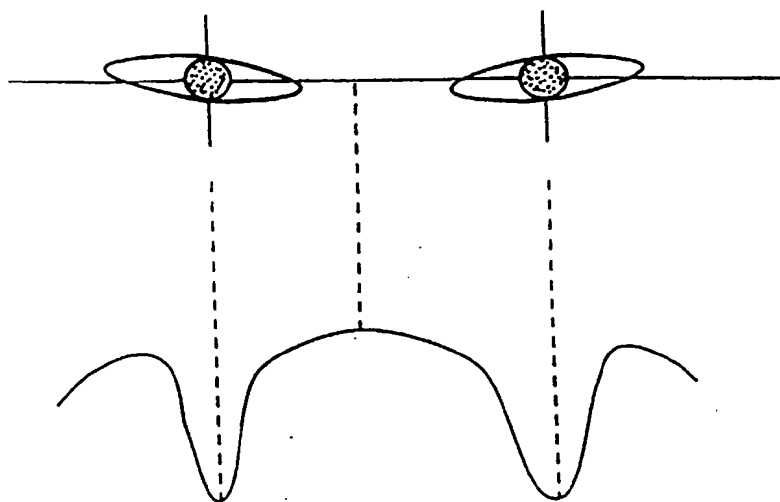
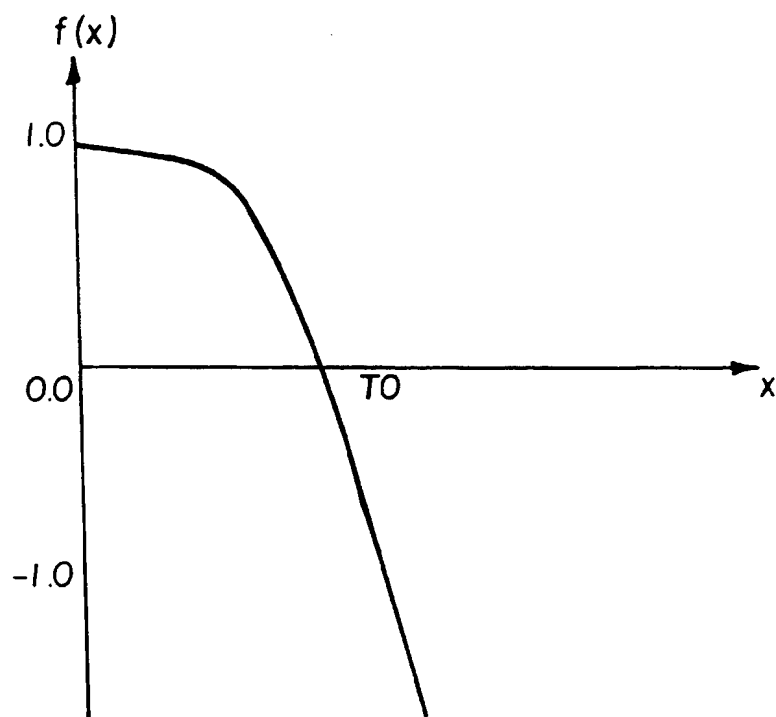


FIG. 8



A PREFERRED SCORING FUNCTION  $f(x)$

FIG. 9

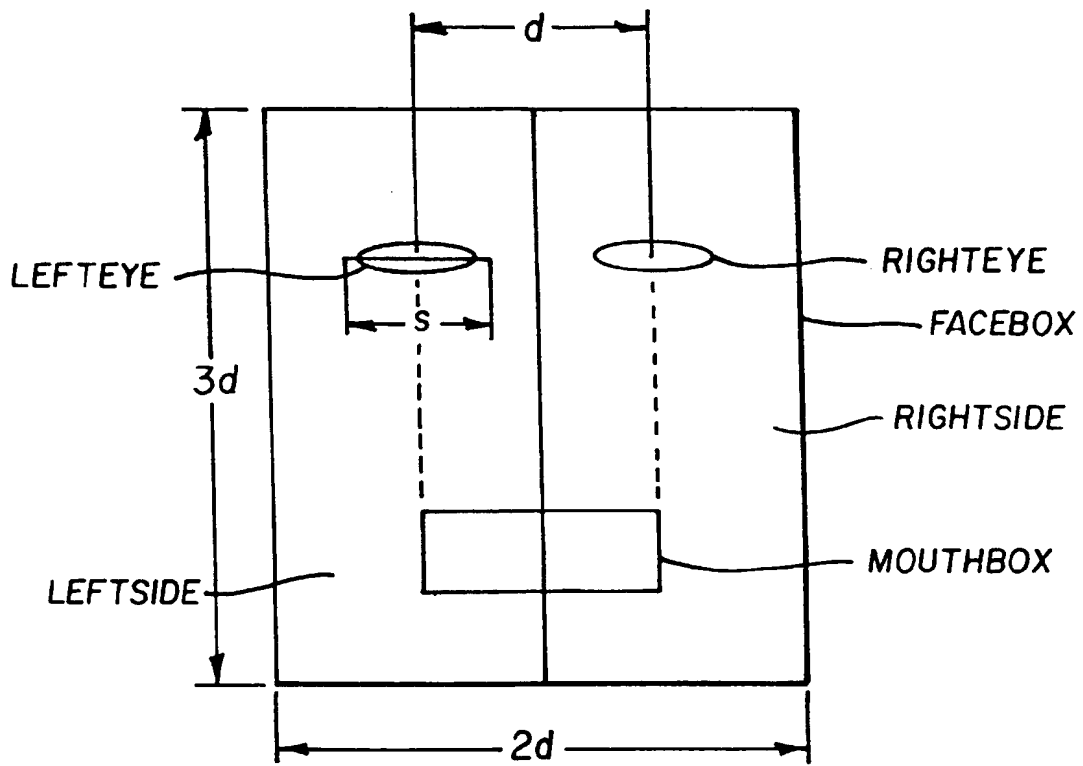
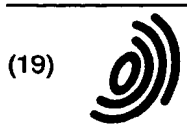


FIG. 10



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 0 899 680 A3**

(12) **EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:  
**12.04.2000 Bulletin 2000/15**

(51) Int. Cl.<sup>7</sup>: **G06K 9/46**

(43) Date of publication A2:  
**03.03.1999 Bulletin 1999/09**

(21) Application number: **98202738.5**

(22) Date of filing: **15.08.1998**

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**  
Designated Extension States:  
**AL LT LV MK RO SI**

(30) Priority: **29.08.1997 US 920844**

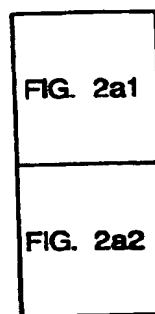
(71) Applicant: **EASTMAN KODAK COMPANY**  
**Rochester, New York 14650 (US)**

(72) Inventor:  
**Luo, Jiebo**  
**c/o Eastman Kodak Company**  
**Rochester, New York 14650-2201 (US)**

(74) Representative:  
**Lewandowsky, Klaus, Dipl.-Ing. et al**  
**Kodak Aktiengesellschaft,**  
**Patentabteilung**  
**70323 Stuttgart (DE)**

(54) **Method for automatic detection of human eyes in digital images**

(57) A computer program product for locating first and second objects, each having substantially the same physical characteristics, and the ratio of the distance between the first and second objects and the size of each object is substantially invariant, the computer program product comprises: a computer readable storage medium having a computer program stored thereon for performing the steps of determining potential flesh regions in an image; determining valley regions in an image; performing template matching for determining a plurality of locations that give a desirable match of the object relative to the template; and performing verification for determining the likelihood of pairs of potential eye candidates at the locations determined in the performing template matching step.



**FIG. 2a**

**EP 0 899 680 A3**



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 98 20 2738

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.8)
Y	CHOW G ET AL: "TOWARDS A SYSTEM FOR AUTOMATIC FACIAL FEATURE DETECTION" PATTERN RECOGNITION,US,PERGAMON PRESS INC. ELMSFORD, N.Y, vol. 26, no. 12, 1 December 1993 (1993-12-01), pages 1739-1755, XP000420368 ISSN: 0031-3203 * chapters 2, 3 *	1-10	606K9/46
Y	SABER E ET AL: "Face detection and facial feature extraction using color, shape and symmetry-based cost functions" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 25 August 1996 (1996-08-25), XP002097369 * the whole document *	1-10	
A	SOBOTTKA K ET AL: "FACE LOCALIZATION AND FACIAL FEATURE EXTRACTION BASED ON SHAPE AND COLOR INFORMATION" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP),US,NEW YORK, IEEE,1996, pages 483-486, XP000704075 ISBN: 0-7803-3259-8 * the whole document *	1-10	
			TECHNICAL FIELDS SEARCHED (Int.Cl.8)
			606K
The present search report has been drawn up for all claims			
Place of search <b>THE HAGUE</b>		Date of completion of the search <b>16 February 2000</b>	Examiner <b>Sonius, M</b>
<p><b>CATEGORY OF CITED DOCUMENTS</b></p> <p>X : particularly relevant if taken alone  Y : particularly relevant if combined with another document of the same category  A : technological background  O : non-written disclosure  P : intermediate document</p> <p>T : theory or principle underlying the invention  E : earlier patent document, but published on, or after the filing date  D : document cited in the application  L : document cited for other reasons  &amp; : member of the same patent family, corresponding document</p>			

EP FORM 1503 03/02 (P0401)